Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference

William J Murphy[1] & Stephen J O'Brien[2]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843-4458, USA. [2]Laboratory of Genomic Diversity, National Cancer Institute-Frederick, Frederick, Maryland 21702, USA. Correspondence should be addressed to W.J.M. (wmurphy@cvm.tamu.edu) or S.J.O. (obrien@ncifcrf.gov).

Published online 15 November 2007; doi:10.1038/nprot.2007.429

Here we describe protocols for designing, optimizing and implementing conserved anchor primers for use in genome mapping or phylogenetic applications, with particular emphasis on homologous gene sequences among mammals. The increasing number of whole genome sequences in public databases makes this approach applicable across a wide range of taxa. Genome sequences from representatives of two or more divergent subclades within a taxonomic group of interest are used to identify candidate local alignments (i.e., exons, exons spanning introns or conserved 5′- or 3′-untranslated regions) that contain sequences with appropriate variability for the chosen downstream application. PCR primers are designed to maximize amplification success across a broad range of taxa, and are optimized under a touchdown thermocycling protocol. Based on the initial optimization results, primers are selected for application in a diverse sampling of species, or for mapping the genome of a target species of interest. We discuss factors that have to be considered for experimental design of broad-scope phylogenetic studies. With this protocol, primers can be designed, optimized and implemented within as little as 1–2 weeks.

INTRODUCTION

The original concept of comparative anchor-tagged genome sequences was developed in the early 1990s, when the comparative mapping community nominated orthologous gene loci spaced across the human genome that would be targeted as syntenic markers in other mammalian genomes[1]. The idea was to align orthologous DNA sequences available from divergent mammalian species in different orders (e.g., human from the order Primates and mouse from the order Rodentia) to identify highly conserved regions from which PCR primers would be designed in the hope that they would amplify the homologous region in species from other distinct mammalian orders. Various PCR-based versions of this approach were formalized in the mid- to late 1990s (see refs. 2–4). In one study, PCR primers from 314 gene loci, termed comparative anchor tagged sequences (CATS), were designed and optimized for use in comparative gene mapping projects[3]. Several other studies suggested similar sets of conserved primers (TOASTS, traced orthologous amplified sequence tagged sites; UMPs, universal mammalian primers)[2,4]. The general theme of these studies was to design primers in closely spaced conserved exons (<2 kb, where known), ideally flanking an intron, which would contain sequence variation adequate for orthologous gene detection, mapping and sequencing in diverse, less-studied mammal species. In the following years, comparative gene maps from several species, based on linkage and/or radiation hybrid approaches, successfully incorporated many of these kinds of markers, and established a firm basis for comparative genome inference across the mammalian radiations[5–9].

Gene mapping projects have been particularly useful for determining ordered orthologous chromosomal segments across species, for positional cloning efforts and for assembly of whole genome shotgun-derived sequence contigs and scaffolds[10]. Early efforts in mapping domestic and companion animal genomes relied on gene-based primers, usually designed from one or two existing genomes (human and mouse) or gene collections in public databases. Today, the number of vertebrate genome sequencing projects has grown to encompass more than 30 genomes, ranging from low-coverage survey (i.e., 2-fold coverage) to finished-sequence quality, deposited and annotated in several public databases. This increase in available genome sequence alignments and phylogenomic tools for multispecies comparisons[11] has given rise to a new wave of model organisms, empowered by well-established genomic resources. These abundant genomes provide access to thousands of potential sequences for PCR primer-based marker design. Currently, however, these tools are available for a very small number of species of biomedical or economic importance. Gene markers derived from these sequenced genomes may be valuable for closely related, 'genome-enabled'[12] taxa, but have more limited utility in species outside 5–10 million years of divergence time. Therefore, full-scale mapping projects for non-traditional model organisms, or targeted sequencing of candidate gene regions in most other species, must utilize sequence information from phylogenetically close relatives. The design of conserved CATS-type primers has broad utility in several applications, including mapping of species that do not enjoy highly developed genomic resources, and targeted amplification of candidate genomic regions across divergent taxa. In addition, the concept and design of CATS-type primers have also been applied and shown to be invaluable as molecular markers in population genetics and deeper family- and ordinal-level phylogenetic studies[13–16]. Here we describe protocols (see **Fig. 1** for an overview) that proved successful for both comparative and phylogenetic inferences, and discuss critical details that we discerned in search for, discovering and applying comparative sequence tag primers in gene mapping and phylogenetic investigations[8,9,13–16]. This approach is only limited by the ability to obtain genomic alignments from a diverse set of two or more species as representatives from the target study group.

Experimental design

Quality assessment of genomic DNA. High-quality genomic DNA is an important factor in PCR success rate, particularly for longer amplicons (>1 kb). DNA can be extracted using a number of methods, such as phenol–chloroform, although this approach is more labor-intensive and produces hazardous material waste. We prefer using commercial kits (e.g., Qiagen DNeasy blood and tissue kit, Qiagen FlexiGene DNA kit), as they are quick and yield more than adequate amounts of highly pure DNA. Genomic DNA stocks should be checked routinely for signs of degradation by running ~200–500 ng on a 1% (wt/vol) agarose gel. A majority of the DNA should be of high molecular weight with little smearing (degradation), extending below 1 kb. Even DNA samples with some fraction of high-molecular weight DNA, but still showing extensive smearing, may not amplify as reliably as non-degraded DNA. DNA should also be evaluated by spectrophotometric quantification. Relatively pure DNA preparations should have an OD_{260}/OD_{280} ratio in the range of 1.7–2.0. Lower values indicate protein contamination, whereas higher ratios indicate the presence of RNA in the sample. Although the latter is of less concern, the former may result in PCR inhibition.

Primer design. We strongly recommend that primers be designed using a computer program that is able to evaluate multiple primer characteristics, such as melting temperature (T_m), GC content, secondary structure formation and primer-end self-complementarity. In our experience, manually designed primers show markedly lower rates of success. There are several primer design programs available freely via the internet. Note that different primer programs calculate T_m with different methods, such that primers can have T_m values as different as 5–8 °C, despite being designed with the same target T_m. We prefer Primer3 (http://fokker.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (see ref. 17). Primer3 provides a fairly simple graphical interface that allows the user to paste in an input sequence, delimit target regions and exclude specific sequences, and modify other parameters such as primer length, GC content and annealing temperature. The default output provides five candidate primer pairs, although a much larger list can be retrieved by simply increasing the output number settings. In addition, the source code may be downloaded for high-throughput generation of primers from many target sequences simultaneously. We recommend that users do not modify the default parameters unless no suitable primers can be found. The advantage of this is that if multiple primer pairs are being evaluated in the same panel of species at a time, they are more likely to be optimized under a similar set of conditions. This will increase the ease and efficiency of most downstream applications, such as PCR purification and sequencing in 96-well formats, reducing setup time and human error.

PCR conditions and primer optimization. Although well-designed PCR primers may work well under a range of PCR conditions, it is advantageous to attempt optimization under as few conditions as possible to facilitate simultaneous amplification of multiple primer pairs in a single 96-well plate. Although gradient temperature blocks are useful for identifying the best conditions for a particular primer pair across a range of magnesium chloride concentrations[5], this often requires expenditure of a large amount of genomic DNA (which may be undesirable for rare or limited samples) and PCR reagents for each primer. A better alternative is to use a touchdown (or stepdown) PCR profile, where the initial annealing steps begin at a relatively stringent starting temperature (60–65 °C), which is lowered every one or two cycles until it reaches a fairly relaxed temperature (50 °C). This allows each primer pair to be optimized in a single tube for a particular magnesium concentration. We generally start with PCR cocktails containing 1.5 and 2.0 mM $MgCl_2$. In some cases, increasing the $MgCl_2$ concentration above 2.0 mM (i.e., 2.5–3.0 mM) will improve amplification yield, but may also introduce non-target bands.

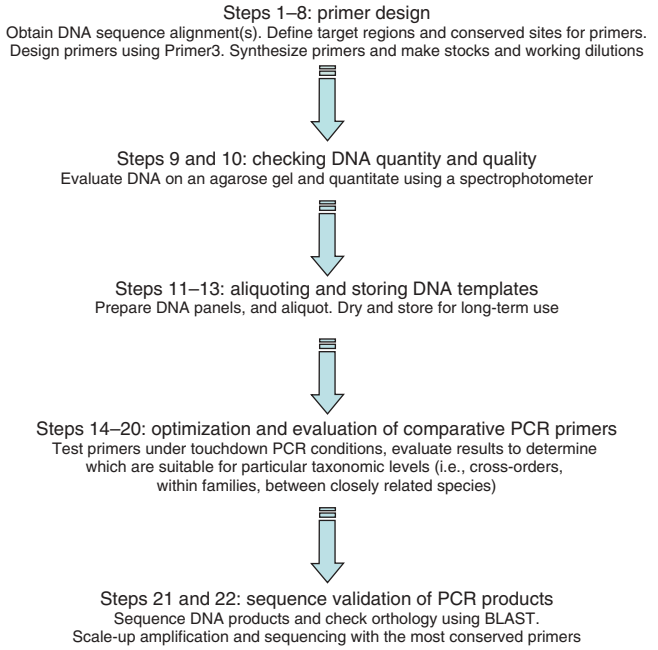

Figure 1 | General flow diagram of comparative primer design, testing and optimization.

Alternative procedures for purifying suboptimal PCR products. Primer pairs that work across the majority of DNA samples will often fail to produce robust amplicons or produce spurious bands in a subset of species. Although cloning of PCR products can be used to isolate a pure product of sufficient DNA quantity, this is time consuming and costly for just one or a few PCR products. One alternative for multibanded products is to excise the desired product from the gel and purify it with a silica or spin-column-based procedure. The other option is to perform a touch-prep, whereby a pipette tip is used to stab ('touch') the appropriate product in an agarose gel. The tip is then vigorously swirled/washed in a small volume (10 µl) of sterile pre-PCR water. A 1–2 µl volume of this template is then used in a subsequent PCR reaction with the same primers (that are now a perfect match with the template) under stringent conditions (i.e., 1.5 mM $MgCl_2$, 60–62 °C annealing temperature for 30–35 cycles).

Factors to be considered when setting up amplifications for downstream sequencing and analysis. When amplifying and sequencing products using several optimized primer pairs across a broad number of species, creating panels of high-quality genomic DNA increases the consistency, speed and efficiency of downstream handling steps. Dilute the DNAs to the same concentration (5–10 ng μl^{-1}) for PCR setup in sets that form multiples of 8 or 12, to accommodate efficient replicate-aliquoting into 96-well plates.

Aliquot these (a total of approximately 25 ng per well) at one time into multiple plates, and dry overnight at room temperature (~21–27 °C), either by covering with clean tissue on an isolated benchtop or in a fume hood. This approach ensures that consistent amounts of template DNA are placed in the same well for each amplification.

Laboratory setup. It is ideal to have separate pre-PCR and post-PCR laboratory rooms (adjoining if possible) or spaces, particularly when dealing with conserved primers and multiple DNAs of related species. The opportunity for contamination of genomic DNA stocks and PCR reagents is a real threat if they are not isolated from equipment that is used to manipulate or handle post-PCR products, where target DNA regions are amplified millions of times in excess of standard genomic DNA concentrations. The pre-PCR lab (or lab area) should be used for DNA aliquoting/pipetting and stock storage, PCR reagent setup and storage, PCR experiment setup, and should contain all the equipment (i.e., centrifuges, vortexes and pipettors) that are required for these procedures. DNA thermocyclers, pipettors for handling PCR products, gel electrophoresis equipment, refrigerator and freezer storage for post-PCR products should all be kept in an isolated post-PCR area far from traffic in the pre-PCR area. Care should be taken to change gloves when leaving the post-PCR area, and that equipment is not moved in between these separate labs/areas. Gel preparation can be performed in the pre-PCR area as long as glasswares are immediately washed and are never taken to the post-PCR area. Gel documentation should be performed in the post-PCR area. Pre-PCR equipment, such as 96-well rubber mats/covers and PCR trays, that are taken to the post-PCR area should be soaked in a 5% (vol/vol) solution of Clorox bleach and rinsed with water before being returned to the pre-PCR laboratory area.

MATERIALS

REAGENTS

- Oligonucleotides (Invitrogen); desalted and rehydrated in TLE buffer (see REAGENT SETUP)
- Platinum *Taq* DNA polymerase, 10× buffer and 50 mM $MgCl_2$ (Invitrogen, cat. no. 10966)
- dNTPs (dATP, dCTP, dGTP, dTTP), 100 mM blend (Applied Biosystems, cat. no. N8080261)
- SeaKem LE high-melting temperature agarose (Cambrex, cat. no. 50004)
- Sterile, molecular biology grade water (Sigma, cat. no. W4502)
- Tris (Sigma, cat. no. T1503) ! CAUTION Irritant. Wear gloves and suitable protective clothing.
- Boric acid (Sigma, cat. no. B6768)
- EDTA disodium salt (Sigma, cat. no. E5134) ! CAUTION Irritant. Wear gloves and suitable protective clothing.
- Bromophenol blue (Sigma, cat. no. B0126)
- Xylene cyanol (Sigma, cat. no. X4126)
- Glycerol (Sigma, cat. no. G5516)
- Ethidium bromide (EtBr; Sigma, cat. no. E2515) ! CAUTION Harmful, carcinogenic. Wear gloves and suitable protective clothing.
- 1 kb-plus DNA ladder (Invitrogen, cat. no. 10787)

EQUIPMENT

- Personal computer with internet access
- Two sets of manual pipettors (Eppendorf), one each for pre- and post-PCR analysis
- Aerosol barrier pipette tips
- Standard microfuge tubes (0.6 and 1.5 ml)
- PCR strip tubes with attached caps (8 or 12)
- 96-well PCR plates and covers
- 96-well plate holders/base (e.g., Applied Biosystems, cat. no. N8010531)
- Benchtop centrifuge with 96-well plate holding rotor (Eppendorf 5804)
- Automatic multichannel repeat pipettors (Matrix) required for pre- and post-PCR analyses
- Reagent reservoirs (Matrix)
- DNA themocycler (Applied Biosystems, 9700)
- Plate mixer for 96-well plates (VWR)
- Manual vortex (VWR), one each for pre- and post-PCR analysis
- Agarose gel electrophoresis apparatus with 24–48 tooth capacity (Owl Scientific)
- Power supply for electrophoresis (VWR)
- Gel imaging and documentation system (Biodoc-It)

REAGENT SETUP

Tris buffers (1 M Tris-HCl (pH 8.0) and 1 M Tris-HCl (pH 7.5)) Dissolve 121 g Tris in 800 ml of milli-Q grade water. Add concentrated HCl to reach the desired pH and make up to 1 liter with water. Autoclave and store at room temperature.

0.5 M EDTA (pH 8.0) Dissolve 93 g EDTA in 850 ml of milli-Q grade water. Add NaOH pellets (~10 g) to bring the solution to pH 8.0 and make up to 1 liter with water. Autoclave and store at room temperature.

TLE buffer $T_{10}E_{0.1}$ buffer, L = low EDTA concentration, as opposed to standard $T_{10}E_{1.}$ Mix 1 ml of 1 M Tris-HCl (pH 7.5) with 20 µl of 0.5 M EDTA (pH 8.0). Make up to 100 ml with water. Autoclave and store at room temperature.

10× TBE buffer for agarose gel electrophoresis Dissolve 108 g Tris and 55 g boric acid in 600 ml of milli-Q grade water. Add 40 ml of 0.5 M EDTA (pH 8.0) and make up to 1 liter with water.

1% EtBr stock solution Dissolve 100 mg EtBr in 10 ml of water. Cover the outside of the bottle with aluminum foil to protect from light and label accordingly.

1% (wt/vol) agarose gel Add 6 g agarose to 600 ml of 0.5× TBE buffer in a 1 liter Erlenmeyer flask. Let it stand for 5–10 min to allow the agarose powder to absorb the buffer; this will prevent boiling over in the microwave. Microwave is kept on highest setting for 4 min, stopped to stir every minute, until agarose is completely melted. Cool in lukewarm water bath until temperature reaches <55 °C. Add 15 µl EtBr stock solution to agarose, swirl and pour into casting tray (~5 mm thick) with appropriate combs. The amount of EtBr may need adjustment. Allow the gel to solidify at room temperature (~30–45 min). ▲ CRITICAL Cool in lukewarm water bath until the temperature reaches <55 °C to avoid warping the gel casting tray.

6× gel loading dye Place 35 ml sterile milli-Q water in a 50 ml conical centrifuge tube (Corning). Add 15 ml glycerol, 0.125 g bromophenol blue and 0.125 g xylene cyanol. Vortex and store at 4 °C. It is useful to make separate gel loading buffers with only one dye (bromophenol blue or xylene cyanol), to avoid dye-based interference/dampening of bands in the gel. For visualizing small PCR products (<2 kb), use a loading buffer with xylene cyanol only ('upper dye buffer'). For visualizing genomic DNA and larger PCR products (>2 kb), use a buffer with bromophenol blue only ('lower dye buffer'). Bromophenol blue and xylene cyanol co-migrate with 300 bp and 2.4 kb DNA fragments, respectively.

1 kb-plus ladder loading standard Mix 5 µl of 1 µg $µl^{-1}$ one kilobase-plus ladder stock solution (Invitrogen) with 18 µl of 6× gel loading dye and 77 µl TLE. Vortex and store at 4 °C for up to 1 month.

PROCEDURE

Comparative primer design and rehydration ● TIMING Approximately 10 min per primer pair

1| Obtain a multiple sequence alignment for a particular genomic segment from two or more species within your group of interest (see **Box 1** for details of how to obtain suitable sequences). The alignment can be done with any of the various alignment packages, such as Clustal (http://bioinf.ucd.ie/software/clustal.html), Se-Al (http://tree.bio.ed.ac.uk/software/seal/)

BOX 1 | RESOURCES FOR IDENTIFYING GENES/GENE ALIGNMENTS

There are many ways to obtain a set of orthologous sequences or potential lists of genes suitable for phylogenomic analysis. A sample of some useful sites and references are listed below.

1. *UCSC Genome Browser* (http://genome.ucsc.edu/). This site contains genome assemblies and multispecies alignments for many vertebrate species.
2. *BLAST* (http://130.14.29.110/BLAST/). Basic local alignment tool is used to find regions of similarity between input DNA sequences. A large number of specific databases can be searched, including whole genome sequence assemblies, expressed sequence tags and trace archives (see below).
3. *BLAST Trace archives* (http://130.14.29.110/BLAST/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&BLAST_SPEC=TraceArchive&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch). This is an excellent repository of raw DNA sequence reads generated by the major genome sequencing centers that can be downloaded for analysis. This sequence database includes a substantial fraction of sequences that are not incorporated into initial genome assembly contigs. These can be searched by BLAST using a reference input sequence to obtain potentially orthologous gene alignments from related taxa.
4. *HOVERGEN database* (http://pbil.univ-lyon1.fr/databases/hovergen.html). A database of homologous vertebrate gene sets that can also be used to visualize multiple alignments and phylogenetic trees.
5. *HOMOLOGENE database* (http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene). An NCBI database for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes. The search query is a gene name or symbol. Information for each gene includes various evolutionary parameters derived from pairwise alignments, conserved domains and links to other gene annotation databases.
6. *Published PCR primers*.

Refs. 2–4, 6, 9 and 13 contain pairs of primers that amplify orthologous sequences in mammals. Also see **Supplementary Table 1**.

and Sequencher (http://www.sequencher.com). Alternatively, one may obtain an alignment of a particular genomic location using the UCSC Genome Browser. It is advantageous to include an out-group species to help identify the most constrained regions of sequence variability for primer design.

2| Identify two conserved stretches (length = 25–30 bp) of sequence with 85–100% sequence identity in the in-group species, flanking a region of sequence variation appropriate for your application (**Table 1**). These are your candidate primer regions.

3| Cut and paste the sequence of one of your in-group sequences into the Primer3 sequence input box (see **Fig. 2** for an example input).

4| Specify the target region to be amplified immediately inside of the candidate primer regions, using either bracket symbols (i.e., [and]) in the input sequence, or by specifying the target region numerically (e.g., 250,200 defines the sequence between bases 250 and 450 as the target PCR region; see **Fig. 2**). For primer pairs that span introns, allow at least 50 bp of exonic sequence internal to each primer. This will ensure that sufficient conserved sequence remains for DNA sequencing-based verification of orthology of the target locus, using BLAST analysis, to related genomes.

5| Exclude all nucleotides before the 5′- (upstream) and after the 3′- (downstream) candidate primer regions using '<' and '>' symbols in the input sequence, or by specifying the excluded regions numerically (e.g., 1,49 means the sequence between bases 1 and 50 is excluded; see **Fig. 2**).

6| Push the 'Pick Primers' button below the sequence input box in Primer3 to produce a list of candidate primers. Primer3 will produce a list of as many as five candidate primer pairs by default (although this list can be increased by changing the 'Number to Return' setting in the Primer3 menu page). Choose primers within the candidate regions that have no sequence mismatches near the 3′- ends in your multispecies alignment; limit any mismatches to the central regions of the primers. If your candidate primer regions contain a few base-pair substitutions between the in-group species, choose the sequence from the species with

TABLE 1 | Genomic regions and their phylogenomic applicability.

Level of divergence	Example	Ideal nuclear gene marker
Very deep	Across vertebrates	Exonic regions of housekeeping genes
Deep	Across mammals	Most exons, some UTRs, UCEs
Moderate	Between orders	Most exons and UTRs, some introns
Shallow	Within orders	Rapidly evolving exons, most introns
Recent	Between species	Long introns (>1 kb), neutral intergenic DNA

UCEs, ultraconserved elements; UTRs, untranslated regions

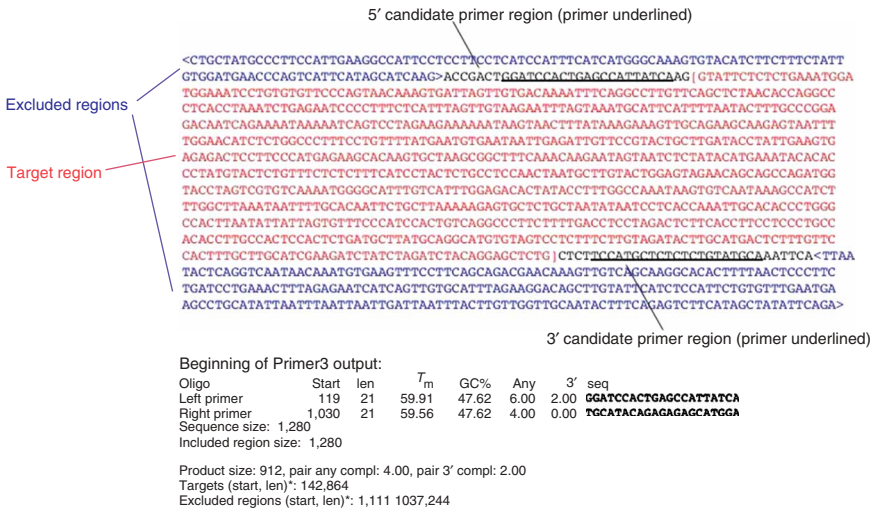

Figure 2 | Example of an annotated DNA sequence used as input for Primer3. Regions to exclude from primer consideration are bounded by '<, >' marks (highlighted in blue). The target amplification region is demarcated by brackets '[,]' (highlighted in red). The remaining regions (in black) are the candidate primer regions identified by the multiple species alignment from your group of interest (see **Box 2**). An example of the beginning output from Primer3 is shown in the figure. The first candidate primer pair is shown (written 5′ to 3′), along with primer lengths, T_m, GC content and other parameters, such as product size. 'Any' refers to the self-complementarity score (whether the primer will anneal with itself or for secondary structure), whereas 3′ refers to the 3′ self-complementarity score (the tendency of the primers to form primer dimers).

the greatest ability to mispair as the input for Primer3 (see **Box 2**). Alternatively, after selecting primers, incorporate degenerate IUPAC-defined sites into your primer sequences (e.g., R = A or G, Y = C or T; see **Box 2** for primer design guidelines).

7| Synthesize primers from a commercial vendor (e.g., Invitrogen, Operon) or by using an in-house DNA synthesizer, specifying a standard desalted and lyophilized product.

8| Rehydrate primers with an appropriate volume of TLE to make a 200 µM stock solution and vortex thoroughly for 15 s. Make a 1:10 dilution in sterile milli-Q water; this will serve as a working dilution (20 µM) for PCR.

Checking DNA quality and quantity ● TIMING Approximately 1 h

9| Run a small aliquot (1–5 µl) of DNA on a 1% (wt/vol) agarose minigel in 0.5× TBE running buffer at 105 V for 15–20 min. If the DNA is of high concentration/viscosity, dilute the sample 3–5-fold in TLE, add an appropriate amount of 6× 'lower'

BOX 2 | GUIDELINES FOR DESIGNING NON-DEGENERATE COMPARATIVE PCR PRIMERS

There are two general approaches for designing primers that are useful across a broad selection of species. The first approach is the design of degenerate primers. In this approach, the primer sequence includes ambiguous bases that follow the IUPAC ambiguity code (e.g., C or T = Y; A or G = R). In practice, degenerate primers work very well, assuming that the alignment from which the primers were designed includes a broad representation of taxa from your study group of interest. If this is the case, then, in theory, there should be a primer synthesized in the mixture that will perfectly match your target sequence. The drawback is that only a fraction of a highly degenerate primer synthesis (e.g., >3 degenerate sites in the primer) will contain a perfect match to the target sequence. If both primers are degenerate, this match fraction is even lower. This can lead to low amplification yields, asymmetric amplification, spurious products, or weak or no DNA sequencing results. Primer concentrations may be increased to improve the results in both PCR and DNA sequencing.

An alternative to designing degenerate primers is to take into account nucleotide mismatches that form weak or partial bonds[18]. For example, G normally binds to C, but will also form base pairs with T. In the example shown below, an alternative to the degenerate primer is to use the T and G at positions 5 and 16 in the primer sequence. The T (human, mouse and opossum) will base-pair with the G (complementary strand to C) in dog and cow. Similarly, the G (human and dog) at position 16 will bind well with the T (complementary strand to A) in mouse, cow and opossum. When mismatches are incorporated, it is best if they are placed in the middle or toward the 5′ end of the primer. Mismatches near the 3′ end should be avoided, particularly A:G and C:C mismatches, which can decrease yield by 100-fold[18]. Because primers containing mismatches have a lower affinity for the target sequence, the total number of mismatches should be kept at a minimum (ideally less than three). One advantage of this approach is that it eliminates the potential problems for DNA sequencing that exist for degenerate primers because following amplification, the same primers are a perfect match for priming the DNA sequencing reaction.

Examples of multispecies alignment

Human	AAGGGTTTCCAAGAGATCACCCCACCAGAAAAGGGTAGGAATGAGCAAGTTGGGAATTTTAGAC
Mouse	AAGGATTTCCAAGAGATCACCCCACCAAAAAAGGGTAGGAATGAGCTAGTTGGGAAGTTTAGAC
Dog	AAGGGTTCCCAAGAGACCACCCCACCAGAAAAGGGTAGGAATGAGCAAGTTGGGAATTTTAGAC
Cow	AAGGGTTTCCAAGAGACCACCCCACCAAAAAAGGGTAGGAATGAGCAAGTTGGGAATTTTAGAC
Opossum	AAGGGTTACCAAGAGATCACCCCAGCAAAAAAGGGTAGGAATCAGCGAGTTGGGAATTATGGAC
Degenerate primer	GAGAYCACCCCACCARAAAA >>forward
Mismatch primer	GAGATCACCCCACCAGAAAA >>forward

loading dye and pipette up and down several times to ensure even mixing.
▲ CRITICAL STEP This will allow the DNA to migrate out of the well evenly and avoid streaking from the well, which would make quantification difficult (see Step 10). Add 1–5 µl of EtBr (depending on the size of the gel and the chamber) to the lower running buffer chamber and swirl to ensure the gel stains evenly. Intact genomic DNA will be present as an intense high-molecular-weight band. Poor quality DNA will appear as a lower molecular weight smear.
? TROUBLESHOOTING

10| To quantify your DNA you may run an aliquot of a high-molecular weight DNA standard of known concentration alongside your samples. For comparison, check your DNA concentration with a spectrophotometer or a nanodrop device, taking both OD_{260} and OD_{280} measurements to determine whether your DNA has RNA or protein contamination.

Aliquoting and storing DNA template dilutions for PCR ● TIMING Approximately 16 h

11| Once you have determined which of your DNAs are of high quality, pick a representative set of DNAs from diverse species in your group under study. Make 5 ng μl^{-1} dilutions in TLE, and assemble the amount required for all amplifications in either 8- or 12-strip 0.2 ml PCR strip tubes.
▲ CRITICAL STEP Always use aerosol barrier pipette tips when handling genomic DNA and PCR reagents. Make sure to include a well with PCR water as template for your negative control.

12| Using an 8- or 12-channel multichannel pipettor, dispense 5 µl aliquots of the diluted DNA into each row or column (depending on the number of samples tested) of a 96-well PCR plate. Lift the plate and visually confirm that each well has received DNA. Cover each 96-well plate with a disposable adhesive plastic cover and briefly (~30s, room temperature) spin down the DNA in a benchtop centrifuge with a 96-well-capacity rotor at ~1,000*g*.

13| Place plates in a stable 96-well format rack/plate base and carefully remove the film to avoid well-to-well splatter. Place the DNA plates in either a laminar air flow hood or on an isolated workbench away from traffic. Cover loosely with large Kimwipes or Saran Wrap and dry overnight. If you are using a laminar flow hood, cover plates with Whatmann 3M paper. The following morning, check the plates to ensure that they are dry; then cover with adhesive plastic covers until ready for use.
■ PAUSE POINT If the plates are not ready for use immediately, they can be stored for up to 6 months at 4 °C, or 1–2 years at −20 °C.

Optimization and evaluation of comparative PCR primers ● TIMING Approximately 5 h

14| Prepare working PCR dilutions (20 µM) of comparative primer stocks. Store at −20 °C.

15| Set up PCR cocktails for each primer pair at both 1.5 and 2.0 mM $MgCl_2$ concentrations (**Table 2**). Vortex well and briefly spin down the contents in a microfuge.

16| Pipette each cocktail into a different well of an 8- or 12-strip 0.2 ml PCR strip tube.

17| Using an automatic multichannel pipettor or a repeat-dispensing pipettor, dispense 25 µl of each PCR cocktail (from the different wells of the PCR strip tube in Step 16) into each set of dried-down DNAs (from Step 13), so that each DNA sample

TABLE 2 | PCR mastermixes.

	Volume to add (µl) for number of samples, *X*			
Compound	***X* = 12 (14)**	***X* = 24 (28)**	***X* = 48 (56)**	***X* = 96 (110)**
Sterile milli-Q filtered water[a]	260.4/256.9	520.8/513.8	1041.6/1027.6	2,046/2018.5
10× PCR buffer	35	70	140	275
50 mM $MgCl_2$ (for 1.5/2.0 mM)	10.5/14	21/28	42/56	82.5/110
dNTP blend (10 mM)	28	56	112	220
Forward primer (20 µM)	7	14	28	55
Reverse primer (20 µM)	7	14	28	55
Platinum *Taq* DNA polymerase	2.1	4.2	8.4	16.5
Final mastermix volume (µl)	350	700	1,400	2,750
Volume to aliquot per well	25	25	25	25

The calculation of the number of samples (X) that will receive mastermix includes enough volume for extra reactions (total shown in parentheses) to account for pipetting error. For example, *X = 12 (14)* indicates that the mastermix contains enough cocktail for 14 tubes but, due to the pipetting errors of typical multichannel pipettors, will safely fill at least 12 tubes. [a]These volumes assume that the mastermix will be added to a PCR plate in which the DNA has been dried down. If using wet DNA, reduce water volume accordingly. In most cases, the reaction volumes may be scaled down to 10 or 15 µl reaction volumes, without any change in amplification results.

receives each different PCR cocktail. Cover the plate with a 96-well cover (rubber mat or adhesive tape that can be used in a thermocycler). Spin down the contents briefly in a centrifuge and allow the DNA to rehydrate for 10 min at room temperature.

18| Rub the plate back and forth on a plate-mixer style vortex for 15 s, avoiding splatter.

19| Spin down the contents briefly and place in a 96-well thermocycler under the following program:

2 min at 95 °C (hot start to activate Platinum-Taq)
2 cycles of 15 s at 95 °C, 30 s at 60 °C, 60 s at 72 °C
2 cycles of 15 s at 95 °C, 30 s at 58 °C, 60 s at 72 °C
2 cycles of 15 s at 95 °C, 30 s at 56 °C, 60 s at 72 °C
2 cycles of 15 s at 95 °C, 30 s at 54 °C, 60 s at 72 °C
2 cycles of 15 s at 95 °C, 30 s at 52 °C, 60 s at 72 °C
30 cycles of 15 s at 95 °C, 30 s at 50 °C, 60 s at 72 °C
5 min final extension at 72 °C

20| Remove the PCR plate from the thermocycler. Bring the PCR plate/tubes to a dedicated post-PCR lab, or an area of the lab that is away from the pre-PCR setup. Use a dedicated post-PCR multichannel pipettor to remove 5 µl of each PCR product and dispense into a clean 96-well plate, already containing 2.5 µl of 3× loading dye per well. Gently vortex the plate to mix sample with dye, and load samples with the post-PCR multichannel pipettor into the wells of a 1% (wt/vol) agarose gel in 0.5× TBE buffer at 105 V for 15–20 min (or longer, depending on product size and resolution desired). Include a DNA ladder (1 kb-plus ladder) for sizing purposes. Products should be clean sharp bands of appropriate size, with minimal smearing or background bands (**Fig. 3**, primer c).

? TROUBLESHOOTING

21| Sequence a subset of the PCR products and confirm orthology using BLAST sequence analysis. Best BLAST hits should be to orthologous sequences from two or more related genomes, and should also hit conserved syntenic blocks in the assembled genomes of more than one species, where available.

22| Based on the initial optimization results, select those primer pairs that amplify at a high success rate (>90%) for testing in a large panel of DNAs. If your plans are to amplify and sequence multiple target loci, aliquot and dry down multiple DNA panels at once, as in Steps 11–13. This will ensure consistency in all downstream applications. Do not include DNAs of poor quality in this panel; instead, amplify these separately as they often produce no/weak amplification products.

● TIMING

Steps 1–8, comparative primer design and rehydration: approximately 2–5 d, depending on the time required for commercial primer synthesis and shipment from vendor.

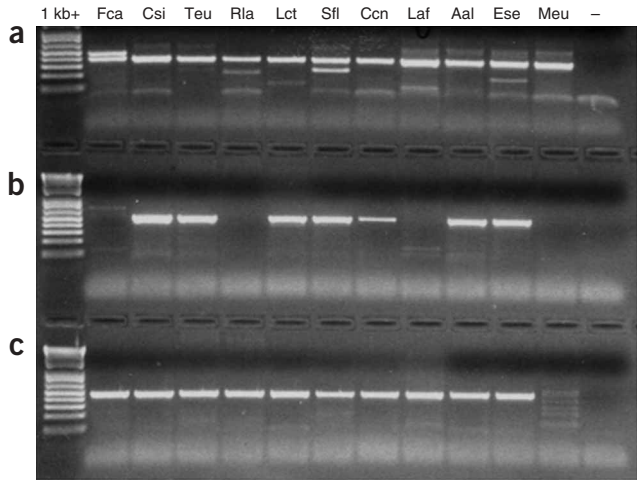

Figure 3 | PCR amplification results of three pairs of PCR primers (**a**, **b** and **c**), each targeting a different nuclear gene locus, tested in 12 mammals from different orders and a water-only negative (–) control. The first lane of each row contains 1 kb-plus DNA ladder. The species IDs/orders are from left to right: Fca (domestic cat, Carnivora); Csi (white rhino, Perissodactyla); Teu (bongo, Cetartiodactyla); Rla (rousette fruit bat, Chiroptera); Lct (ring-tail lemur, Primates); Sfl (rabbit, Lagomorpha); Ccn (beaver, Rodentia); Laf (African elephant, Proboscidea); Aal (European mole, Eulipotyphla); Ese (6-banded armadillo, Cingulata); Meu (tammar wallaby, Diprotodontia). Amplifications were performed in 25 µl reactions at 2.0 mM $MgCl_2$ concentration. The first primer pair (**a**) amplifies a single product of the expected size (~400 bp) in all taxa, but also produces some secondary products in several species. This primer requires further optimization either by lowering magnesium chloride concentration or by increasing the starting annealing temperature. The second primer pair (**b**) amplifies a single product of the expected size (~500 bp), but fails to amplify or produces a band of the incorrect size in four of the species (cat, bat, elephant and wallaby). This primer pair may be suitable for study in some eutherian orders, and may need further optimization within other orders. The third primer pair (**c**) amplifies a single product of the expected size (~450 bp) in all eutherians tested, but produces a multibanded product in marsupials (wallaby). This primer pair would be ideal for using across all eutherian taxa (see **Fig. 4**).

Steps 9 and 10, checking DNA quality and quantity: approximately 1 h
Steps 11–13, aliquoting and storing DNA template: approximately 16 h
Steps 14–22, optimization and evaluation of PCR primers: approximately 5 h

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 3**.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible cause	Solution
9	DNA is degraded	Poor quality of tissue sample	Obtain new material, re-extract
		DNA sample stored for too long or in improper buffer	Re-extract DNA and resuspend in TLE or other Tris-EDTA buffer to stabilize for long-term storage
20	No DNA amplification	DNA is degraded or contains PCR inhibitors	Re-extract DNA to obtain higher quality DNA. If PCR dilutions of DNA are old (>6 months), make new dilutions from stock Make new primers that amplify smaller amplicons (<500 bp) Perform additional DNA purification with filtration devices
		Primers are too different from target sequence	Increase $MgCl_2$ concentration and/or lower annealing temperature (start at 55 °C and step down to 45 °C)
		Error in PCR setup	Redesign primers using sequence from a more closely related species Recheck PCR-mix calculations and age of reagents; repeat PCR
	Faint DNA bands	DNA concentration too low	Increase DNA template concentration Increase the number of PCR cycles to 45 or 50
		Primers are too different from target sequence	See above
	Multibanded products	Primers not specific, amplify paralogs	Re-examine PCR primers. Use relaxed BLAST search to see if they match a single locus or multiple loci in the closest genome assembly available Decrease $MgCl_2$ concentration and/or increase annealing temperature in PCR (start at 65 °C and step down to 60 or 55 °C). Reduce annealing time and/or extension time Redesign primers

ANTICIPATED RESULTS

The success rate of comparative PCR primers will depend on a number of factors, including number and divergence of taxa in your original multispecies alignment, rate of nucleotide substitution across your group of interest (faster evolving species tend to have a higher failure rate) and whether your target gene is part of a gene family with closely related paralogs. In our experience, primers designed from only human–mouse alignments have a final eutherian-wide (i.e. amplifies in >90% of species) success rate of about 15–20% (see ref. 13). For example, in an initial screen of 103 nuclear gene primers, only 20 proved suitable for amplification and sequencing across more than 80% of the taxa, without further optimization, and only 15 of these were incorporated into a final analysis[13]. Most primers have a greater than 90% initial success rate within more restricted groups with less sequence divergence, such as the order Chiroptera (bats)[15] and family Felidae (cats)[16]. We include a list of PCR primers that have amplified consistently in a number of mammalian clades, and may work well or be optimized in other mammalian orders, or even other vertebrate groups (see **Supplementary Table 1** online).

Figure 3 shows a gel loaded with PCR products from three pairs of conserved primers, each targeting a different nuclear gene locus, tested in 11 mammals and a water-only negative control. Depending on how conserved the sequences are, and how many diverse species were used in the DNA alignment, one can expect between 10% and 50% of comparative PCR primers to be broadly amplifiable across a wide range of taxa (e.g., eutherian mammals). Conservation rates within more restricted taxonomic groups (i.e., families, orders) often have higher PCR success rates—greater than 90%. **Figure 4** provides an example of a comparative primer that is nearly universal across eutherian orders. Of the species assayed, 96% produced a single band of predicted size. Typically, a small number of taxa do not amplify or produce multibanded products (particularly species with accelerated rates of nucleotide substitution). Further optimization may be needed to produce a single-banded product in these taxa.

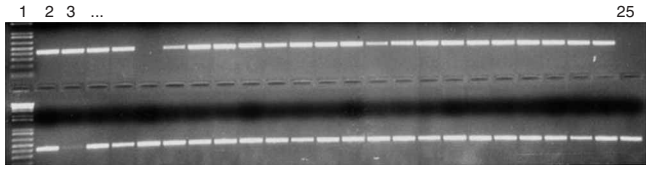

Figure 4 | PCR amplification results for a successful nuclear primer pair (shown in **Fig. 3**, primer c) tested in 47 different eutherian mammal DNAs, suitable for large-scale phylogenetic analysis. The last lane of row 1 (lane 25) is a negative control. The first lane in each row is an aliquot of 1 kb-plus DNA ladder. These results are typical for a well-optimized and conserved nuclear primer pair. DNA sample 2 in row 2 produced a faint product; however, it may still provide sufficient template for DNA sequencing. Note that the DNA sample in lane 6 failed to amplify.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS

We thank Jan Janečka and Alison Wilkerson for helpful comments on this manuscript. This work was funded by the National Institutes of Health and the National Science Foundation.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. O'Brien, S.J. *et al.* Anchored reference loci for comparative genome mapping in mammals. *Nat. Genet.* **3**, 103–112 (1993).
2. Venta, P.J., Brouillette, J.A., Yuzbasiyan-Gurkan, V. & Brewer, G.J. Gene-specific universal mammalian sequence-tagged sites: application to the canine genome. *Biochem. Genet.* **34**, 321–341 (1996).
3. Lyons, L.A. *et al.* Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* **15**, 47–56 (1997).
4. Jiang, Z., Priat, C. & Galibert, F. Traced orthologous amplified sequence tags (TOASTS) and mammalian comparative maps. *Mamm. Genome* **9**, 577–587 (1998).
5. Lyons, L.A., Kehler, J.S. & O'Brien, S.J. Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *J. Hered.* **90**, 15–26 (1999).
6. Murphy, W.J., Shan, S., Pecon-Slattery, J., Chen, Z.Q. & O'Brien, S.J. Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. *Genome Res.* **9**, 1223–1230 (1999).
7. Chen, Z.Q., Lautenberger, J.A., Lyons, L.A., McKenzie, L. & O'Brien, S.J. A human genome map of comparative anchor tagged sequences. *J. Hered.* **90**, 477–484 (1999).
8. Murphy, W.J. *et al.* A radiation hybrid map of the cat genome: implications for comparative mapping. *Genome Res.* **10**, 691–702 (2000).
9. Menotti-Raymond, M. *et al.* Second generation integrated linkage and radiation hybrid maps of the domestic cat. *J. Hered.* **94**, 95–106 (2003).
10. Hitte, C. *et al.* Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. *Nat. Rev. Genet.* **6**, 643–648 (2005).
11. Rosenbloom, K. *et al.* Phylogenomic resources at the UCSC genome browser. in *Methods in Molecular Biology: Phylogenomics* (ed. Murphy, W.J.) (Humana Press, Totowa, NJ, 2007).
12. Kohn, M., Murphy, W.J., Ostrander, E.A. & Wayne, R.K. Genomics and conservation genetics. *Trends Ecol. Evol.* **21**, 629–637 (2006).
13. Murphy, W.J. *et al.* Molecular phylogenetics and the origin of placental mammals. *Nature* **409**, 614–618 (2001).
14. Roca, A.L., Georgiadis, N. & O'Brien, S.J. Cytonuclear genomic dissociation in African elephant species. *Nat. Genet.* **37**, 96–100 (2005).
15. Teeling, E.C. *et al.* A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584 (2005).
16. Johnson, W.E. *et al.* The late Miocene radiation of modern Felidae: a genetic assessment. *Science* **311**, 73–77 (2006).
17. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
18. Palumbi, S. Nucleic acids II: The polymerase chain reaction. in: *Molecular Systematics* 2nd edn. (eds. Hillis, D.M., Moritz, C., Mable, B.K.) (Sinauer, Sundlerland, MA, 1996).